DOCUMENT RESUME

ED 296 578 FL 017 458

AUTHOR Cohen, Andrew D.
TITLE Research in Reading in a Second Language: Discussion Paper. Revised.
PUB DATE Apr 88
NOTE 16p.; Revised version of a paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages (22nd, Chicago, IL, March 8-13, 1988).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Literature Reviews; Measurement Techniques; Models; *Reading Comprehension; *Reading Instruction; *Reading Materials; Reading Processes; *Reading Research; *Reading Tests; Recall (Psychology); *Second Languages

ABSTRACT

Four papers on reading in a second language are reviewed and discussed. They include a study of assessment and the difficulties inherent in the measurement of a construct as complex as reading comprehension. The other three deal with variables facilitating comprehension in reading. One supports an interactive model of reading rather than a skills-based model for promoting reading success. Another suggests that texts providing better contextual and linguistic clues would stimulate better comprehension and recall, and the third proposes training in topical structure analysis as a means for improving comprehension. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

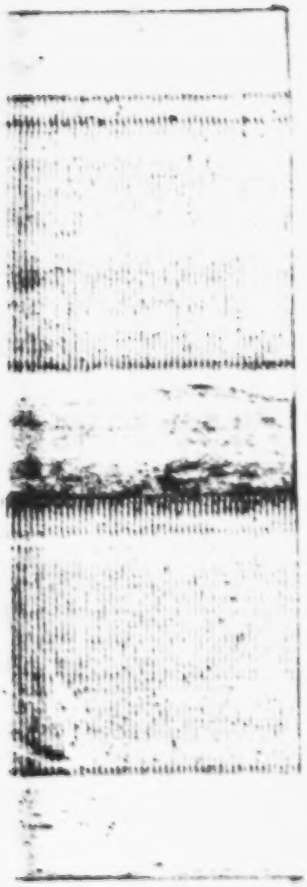
ED296578

TESOL '88 Colloquium on Research in Reading in a Second Language: Discussion Paper

Andrew D. Cohen

Hebrew University of Jerusalem

Revised Version, April 1988



"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. Cohen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL017458

BEST COPY AVAILABLE

Alderson, J. Charles. Testing reading comprehension skills. Lancaster, England: U. of Lancaster.
Devine, Joanne. L2 models of reading and reading strategies. Saratoga Springs, NY: Skidmore College.
Barnitz, John G. & Speaker, Richard B., Jr. 1988. The roles of context and linguistic structure on second language readers' comprehension and retention of poetry. New Orleans: Dept. of Curriculum and Instruction, U. of New Orleans.
John Riggles, Jeff Nagoto, Jim Coady, & Kouider Mokhtari. Improving reading comprehension through topical structural analysis. Athens, OH: Ohio Program in Intensive English, Ohio University.

The first of the four papers I will be discussing, Alderson's, deals with the issue of assessment and the problems inherent in the measurement of so complex a construct as reading comprehension. The other three all deal with variables that facilitate comprehension in reading. Devine claims that those who internalize an interactive model of reading are likely to be more successful readers than those who have internalized a model that is skills-based. Barnitz and Speaker would suggest that texts providing better contextual and linguistic clues will also stimulate better comprehension and recall. Finally, Riggles, Nagoto, Coady, and Mokhtari see training in topical structure analysis as a means for improving comprehension. I will look at each paper in turn, with an emphasis on the raising of issues intended to promote further discussion.

Alderson: Testing Reading Comprehension

The first insight that I draw from this paper is that there may be cognitive hierarchies and implicational scales which are useful for the field of reading assessment, but that we have yet to operationalize these adequately in our language tests. In other words, it would appear that at the present time, we cannot be sure that any given item will assess any given skill or level of skill, nor that the answer on

one item indicates to us how another item will be answered. Hence, my conclusion is that the high/low dichotomy is not very useful.

Perhaps part of the problem is that some of the so-called "enabling skills" are actually multifaceted. In other words, the skill could take on different characteristics depending on the particular reading strategy used to actualize it. Sarig & Folman (1987), for example, present a taxonomy which helps demonstrate that what is ostensibly one reading skill, say, "skimming a portion of text,"[1] could be realized by means of strategies at different taxonomic levels. A given reader could actually be operating on one or all of these levels in performing the "skill." The levels are technical facilitation, clarification and simplification, coherence detection, and metacognitive monitoring. Hence, the psychomotor act of skimming text could reflect technical facilitation, in this case, cutting down the amount of detail taken in and the amount of time expended in reading. Such skimming through a portion of a text could represent a clarification and simplification strategy, assuming the material were deemed redundant or unessential for understanding the essence of the text. The same skimming could reflect coherence detection if the focus of the reader were, say, on how the main idea in each paragraph relates to that of the other paragraphs. Finally, the skimming could also be an instance of metacognitive monitoring if the reader has skimmed, say, after checking for comprehension and realizing that too much time is being spent on this one section of the article with dubious results.

[1] Skimming - a) surveying to obtain the gist, b) scanning for specifics, is #12 on the Weir 1983 list.

Futhermore, it is unrealistic, as Alderson points out, to expect any given test item to elicit only one skill. It may be that several multifaceted skills are being tested. In the case-study portion of his paper, Alderson also notes how the skill that is intended to be tapped may, in fact, not be tapped, and that one or more other skills are actually utilized by the respondent.

Thus, it is helpful to know both what skills an item could potentially test, as well as what skills it actually tests. The problem in trying to obtain this kind of information is that we have to consider the perceptions of:

1) the test constructor, and

2) the test taker.

And if we are concerned about the content validity of our test, as we should be, then we may wish to consider the perceptions of:

3) the outside "experienced teacher" judges, as Alderson and others have done in the current study.

This paper focused among other things, on the discrepancy between the test constructor's and outside judges' perceptions (see Alderson's Table 3). What the paper did not tap are the perceptions of the respondents as to what skills each item was testing. Then of course, the crucial aspect is to determine what skills are <u>actually</u> being tested -- by means of verbal report techniques.

The discrepancy between the test constructor's and outside judges' perceptions as to the skills being tested by various items simply calls attention to the complexity of the "skills" issue. It would appear that the test constructor and the teacher judges had a different sense as to what items testing particular reading skills

look like. Since part of the problem may have been in interpreting what a given skill actually entailed, perhaps a useful exercise would have been to obtain verbal report data from several of the judges regarding the way in which they arrived at their judgments. It also would have been useful to collect the same kind of data from the test constructor as well.

The second part of the paper, featuring a case-study with the two respondents P. and J., is excellent -- the kind of work we need much more of. The fact that little difference was found between introspective and retrospective data is not so surprising. The retrospection involved "immediate retrospection" -- within an hour after the test taking. P.'s experience of amazement was in part a reflection of the benefits of think-aloud and introspective work -- that the informants get a series of insights about the way that they function in a given situation.

The findings speak loudly about problems of reliability and validity in test taking: the difficulty in understanding the test items, the lack of strategies for doing test items (e.g., P.'s just plowing through the text and then reading the questions), the special kind of reading evoked in order to find a word appearing as a multiple-choice alternative, the effects of multiple-choice on finding the correct answers, translation problems (e.g., J.'s answer to #9 was OK in his native language, Spanish, but problematic in English), and a misinterpretation of the instructions (item 3). There are also instances where the respondents, in fact, displayed the expected skills (item 5 -- separating relevant from irrelevant information, item 6 -- comparison of two points of view). In the section that the

items were taken from, respondents were to read a set of questions and then reread a passage that they had presumably already read in order to answer those questions. Suggested means as to how to answer the items were not provided. The text considered relevant in the answering of items was almost always specified.

The conclusions seem quite justified -- that the procedures used for answering test items may not reflect those that the test constructors intended to elicit. It would appear that a test constructor is working on a set of hunches as to what will be called for to correctly respond to an item. In reality, there seem to be various means for obtaining an answer -- whether correct or incorrect -- and with a greater or lesser connection to the skill in question. On the basis of this study, it would appear that:

(1) obtaining a correct response may not require the use of the designated skill.

(2) the use of that skill may not necessarily result in a correct response.

(3) the answering of a given item may call for more than one skill.

(4) an <u>incorrect</u> answer is not necessarily an indication that the respondent did not possess the skill. The problem could simply have been in interpreting how to do the item or in an aversion for that type of item (e.g., item 9, calling for a reading of a graph).

I would reiterate a point which I raised above, namely, that the very same skill could be realized at different levels of cognitive processing. I agree with Alderson that more research is called for.

Devine: Models of Reading and Reading Strategies

Devine posits that there are different internalized models of reading which all readers can describe in themselves and which guide them in their reading. The two types of readers she deals with empirically are the "interactive" readers and the "skills" readers. These categorizations are based both on oral interviews with the readers in which they describe their reading habits and on verbal report protocols obtained from the readers as they read a selected text. Her point is that the interactive readers are more successful than the skills readers at comprehending text.

This study actually builds on two previous studies (Devine 1984, Devine 1987). What Devine calls an "internal model" is actually a composite picture based on what readers say they do and analysis of their verbal report protocols collected while they are in the act of reading. Her categorization of readers into "interactive" and "skills" readers is derived in part from work by Harste & Burke (1977) where approaches to reading are referred to as "meaning-centered," "word-centered," and "sound-centered." With regard to the model, it is worth pointing out that rather than a neat split between interactive and skills readers, there is probably more of a continuum where different readers exhibit both kinds of reading depending on the nature of the source text (its length, complexity, and coherence), the nature of the task (time alloted, specific instructions given), and the background of the reader (schemata, familiarity with topic of text, expertise in reading strategies).

For the purposes of this study, an effort was made to select students at the two ends of the spectrum. Apparently the 14 readers

selected were drawn from an initial pool of 240, from which 26 were selected and then reduced to 14 based on stringent criteria. The pitfall of this kind of sampling is that only extreme cases are being studied, and so the results cannot be said to reflect the patterns of the vast majority of readers. Although the paper does not point this out, none of the subjects had received much specific instruction in reading strategies (Devine, Personal Communication).

Given that reading comprehension involves an interaction between the reader and the specific text with its features, an analysis of the baby talk text for its discourse features would have been helpful. Without it, we can only speculate as to what specific features could have caused problems. It would be useful to know more about the amount of the text that the fourteen readers actually comprehended -- since there can be a discrepancy between learners' impressions/attitudes about how they accomplish reading tasks and their actual performance on the tasks. The readers completed a 15-item multiple choice post test and performed oral retelling. The author reported (Personal Communication) that the retelling was not useful because of variable handling by the research assistants. The issue of what caused difficulties in the text is important since it is possible that a skills-based strategy such as focusing on vocabulary could be more effective than a top-down/interactive one for dealing with certain texts or certain portions of a text.

It is commendable that two reading specialists in Devine's study checked all coding decisions. Mention was not made, however, as to whether there was discrepancy between the investigator's ratings and

those deemed by the specialists to be appropriate. It was also not clear as to whether the specialists agreed with each other.

Whereas it is getting increasingly common for reading strategy studies to include a treatment, in this case the study is non-interventionist, or perhaps, pre-interventionist. The author concludes with the recommendation that teachers find out the kinds of readers present in a given class -- most likely not extremes -- and then on that basis determine the aid that they should get -- presumably aid to skills readers in how to be interactive readers. Of course, it would be possible to argue that interactive readers may also need some training in how to improve their skills in reading -- e.g., their word attack skills.

In sum, the author seems to have found that those predetermined styles were borne out by oral reading performance. My lingering question: Would reading proficiency be consistent with the strategy information in this study? Actually, the study is thin on detail about strategies. Perhaps future studies by the author could provide such detail to help substantiate the claims being made.

Barnitz and Speaker: Context and Linguistic Structure

Barnitz and Speaker do a masterful job of reviewing an enormous amount of literature relating to the nature of the reading process, schemata and context, linguistic structure, and hierarchical effects (11 pages worth!). In their review they give a composite definition of reading: "...a multilevel, interactive, hypothesis generating process in which a reader constructs meaning from a text with a complex orchestration of psycholinguistic processes within a particular

sociocultural situation." Their review reminds us that studies are accumulating in areas which were previously left to speculation -- e.g., the influence of cultural background on reader's ability to recall cultural content in text. They bring us to poetry by indicating that this area has been underinvestigated, thus warranting the current study.

As a methodological aside, we might note that the statistical procedures (e.g., MANOVA) seems rather sophisticated given the small sample -- i.e., 17 seventh graders split into three groups, thus 5-6 per group.

Their first findings was that advanced subjects produced more inferential responses than seventh-grade subjects, suggesting to the authors that inferential text processing increases with language proficiency. Yet the other paper written by the authors on the basis of the same data (Barnitz & Speaker 1987) demonstrated that 7th graders were busy inferencing -- just not inferencing correctly. It would be nice to have some details as to the types of inferences that were made.

Their third finding was that subjects who produced more micropropositions in their recall of text also used more linguistic markers. I wonder whether it would be possible to produce more micropropositions and yet <u>fewer</u> linguistic markers. The poem is short so that the universe of linguistic structures is quite limited. The linguistic markers are tense, plurals, cohesive ties, and agreement. How could a microproposition be specified without transmitting linguistic information of some kind? Thus, the positive correlation between the two is not surprising.

Perhaps this is a minor quibble but the authors refer to "deep and surface structures from the text" as being evident in recall protocols. What the authors mean here are "subordinate and superordinate propositions," also referred to as "micro and macropropositions." This point just reminds us about the problems of terminology. Certainly the terms "deep and surface structure" call up a number of associations totally unrelated to recent work in text structure.

The fourth finding was that advanced subjects recalled more than the other two groups. Also, propositions high in the text hierarchy were recalled better than those at the middle level than those at the low level. The authors' surprise at not finding a context effect is not all that surprising actually. The context here was not essential -- the topic of the text was interpretable without the need for a title. The picture was also not essential.

The researchers are to be commended on their research effort. They have identified and incorporated in their study some of the crucial variables of concern to researchers of L2 reading. But for researchers interested in cognitive processes, not much information is given regarding application to the ESL reading class. We are reminded that it is easier for readers to make inferences in a language they are more proficient in. Finally, as the researchers point out in their other paper on this same poem, this is just one poem so its universe of macro- and micropropositions is limited.

This paper describes two studies on recall and how to improve it.[2] The basic thesis is that "topical structure analysis" (TSA)[3] can provide an important semantic map for getting through dense academic prose. Such analysis is seen to provide the rhetorical structure for understanding and remembering text.

First, there is some unclarity in the sampling procedures. It appears that the control group although matched on the basis of a reading test with experimental students, was nonetheless engaged in more intensive EFL study. Some of the experimental students were, in fact, regular full-time students. Does this mean they already had had the same intensive English? Whatever the case, the experimental group was in a writing course and the control group in a course where one third of the time was devoted to reading. It would be useful to have more information on the nature of their instruction. Were topic sentences and supporting ideas stessed in that course -- perhaps without being given the TSA nomenclature?

Furthermore, the experimental group received only four hours of instruction on the treatment variable -- TSA. Also, due to the lack of verbal report work, even just self-report (wherein readers report what they do generally),[4] it is not known whether the experimental group actually used their training on the short 12-item multiple-choice test that was designed. Also, it is not clear in what ways successful performance on the test was facilitated by TSA. The very

[2] I only had access to the write up of one of the studies so my remarks are limited to that one.

[3] A method of identifying how sentence topics progress from one to another, and how the text as a whole relates to the discourse topic and subtopics; Connor 1987.

[4] Other forms of verbal report include self-observation (introspection and retrospection) and self-revelation (think aloud).

fact that it was in a multiple-choice format meant that, as the authors point out, it may not have been "sensitive enough to reveal the hypothesized effect."

They suggest that a summary would have revealed comprehension and recall more effectively. Actually, that is not necessarily true. My own recent work with summaries (Cohen 1988) would suggest that using summaries as a testing technique is also problematic. A summary does not necessarily reveal accurately the respondent's level of comprehension. Summaries of foreign-language texts may contain material lifted out of the source text with little understanding of its meaning -- especially when the summaries are written in that foreign language. In addition, the summary may be of a shotgun nature -- an excess of detail, within which the correct information is found. On the other hand, the summary may be so general and/or vague, that it is not clear whether it reflects genuine understanding or not. Likewise, the organization of the text itself (e.g., an obvious set of subheadings) may "give away" the topical information. Even if the summary <u>does</u> reveal what the respondent understands about the text, the particular rater may not assess the summary accurately, due to a variety of intervening factors such as problems with the respondent's handwriting, lack of adherence to the scoring key, and so forth.

Finally, it is pointed out that recall of semantic text structure was not assessed. Their second study involved the use of recall and the results were still negative. It is likely that the recall procedure elicited written texts resembling in some ways the same kinds of summaries that are critiqued above for their weaknesses.

As is evident from the above discussion, the four papers addressed here all raise issues of current concern to researchers, administrators, teachers, and learners. There is a keen desire to identify the crucial factors in successful reading comprehension in order to teach reading skills more effectively. There is a commensurate interest in improving the testing of reading comprehension so as to better understand what has been attained and what is still lacking in the skills of readers. Papers such as these will continue to be valuable as long as the issues remain unresolved.

References

Barnitz, J.G. & Speaker, R.B., Jr. 1987. An exploratory study of inferential poetry comprehension: The effects of context of second language learners' inferences. New Orleans, LA: Dept. of Curriculum and Instruction, University of New Orleans.

Cohen, A.D. Forthcoming. The taking and rating of summarizing tasks on a foreign-language reading comprehension test. In C. Hill & K. Parry (Eds.), The test at the gate: towards an ethnography of reading assessment.

Connor, U. & Farmer, M. 1987. The teaching of topical structure analysis as a revision strategy for ESL writers. Indianapolis, IN: English Department, Indiana University.

Devine, J. 1984. ESL readers internatlized models of the rading process. In J. Handscombe et al. (Eds.), On TESOL '83. Washington, D.C.: TESOL, 95-108.

Devine, J. 1987. Models of reading and schema activation. Saratoga Springs, NY: Department of English, Skidmore College.

Harste, J.C. & Burke, C.L. 1977. A new hypothesis for reading teacher research: both teaching and learning of reading are theoretically based. In P.D. Pearson (Ed.), Reading: theory, research and practice. Clemson, SC: The National Reading Association, 32-40.

Sarig, G. & Folman, S. 1987. Metacognitive awareness and theoretical knowledge in coherence production. Ramat-Aviv: Everyman's University.